Assessment of Procedural Skills Using Virtual Simulation Remains a Challenge

Siska Van Bruwaene, PhD,† Marlies P. Schijven, PhD,‡ and Marc Miserez, PhD*§*

*Center for Surgical Technologies, Leuven, Belgium; †Department of Urology, University Hospitals, Leuven, Belgium; ‡Department of Surgery, Academic Medical Center, Amsterdam, The Netherlands; and §Department of Abdominal Surgery, University Hospitals, Leuven, Belgium

OBJECTIVE: The LAP Mentor is a procedural simulator that provides a stepwise training for laparoscopic cholecystectomy. This study addresses its "construct" validity that is present when a simulator is able to discriminate between persons with known differences in performance level on the laparoscopic cholecystectomy in real life.

DESIGN: Three groups with different skill levels performed 2 trials of 4 distinct parts of the cholecystectomy procedure (cholecystectomy exercises) and 1 full procedure on the LAP Mentor. Assessment parameters concerning the quantity and the quality of performance were compared between groups using the Kruskal-Wallis and Mann-Whitney *U* tests.

SETTING: The entire research was performed in the Center for Surgical Technologies, Leuven, Belgium.

PARTICIPANTS: For study purposes, 5 expert abdominal laparoscopists (>100 laparoscopic cholecystectomies performed), 11 surgical residents (10-30 cholecystectomies performed), and 10 novices (minimal laparoscopic experience) were recruited.

RESULTS: With regard to the quantity of performance (time needed and number of movements), the experts showed significantly better results compared with the novices in the cholecystectomy exercises. Only in the full procedure, the results of all the parameters (except speed) were significantly different between the 3 groups, with the best results observed for the experts and worst for the novices. With respect to quality of performance, only the parameter "accuracy rate of dissection" in exercise 3 showed significantly better performance by the experts.

CONCLUSIONS: Only the full procedure of the LAP Mentor procedural simulator has enough discriminative power to claim construct validity. However, the lack of quality control, which is indispensible in the evaluation of procedural skills, makes it currently unsuited for the assessment of procedural laparoscopic skills. The role of the simulator in a training context remains to be elucidated. (J Surg 71:654-661. © 2014 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: laparoscopy, virtual simulation, construct, validation, LAP Mentor, assessment

COMPETENCIES: Patient Care, Professionalism, Practice-Based Learning and Improvement

INTRODUCTION

"See one, do one, teach one" has been the cornerstone of surgical training programs for years. However, with the advent of minimally invasive surgery, technical demands on surgeons have increased, whereas at the same time, training opportunities in the operating theater have diminished because of problems, such as restricted work-hour regulations, legal issues, and time pressure. These changes have created the need for supplementary skills training in a safe laboratory environment where surgical simulation permits learning through "trial and error" without endangering patients' lives. Virtual reality (VR) as a training model is an upcoming tool in these surgical training programs.[1-3]

For basic laparoscopic skills, as well as for more advanced laparoscopic skills, such as suturing and knot tying, the role of the virtual simulators is still under debate. More robust and inexpensive video trainers are both seen as equally effective[3-5] or even superior[6] and apparently more appealing and realistic to the trainees.[4-7] In the more advanced stages of laparoscopy training, focus shifts toward procedural dissection skills. The current training models include live anesthetized animals and animate cadavers that are costly,

The purchase of the LAP Mentor virtual trainer by the Center of Surgical Technologies, Leuven, was partially sponsored by Ethicon Inc (Johnson & Johnson medical). This funding was completely independent from this study and did not influence its results. Drs Van Bruwaene, Schijven, and Miserez have no conflicts of interest or other financial ties to disclose.

Correspondence: Inquiries to Siska Van Bruwaene, MD, PhD, Parijsstraat 1, Vlezenbeek 1602, Belgium; fax: (032) 568-0132; e-mail: siska.vanbruwaene@gmail.com

  1931-7204/$30.00 http://dx.doi.org/10.1016/j.jsurg.2014.01.005

require complex infrastructures, differ from human anatomy, and suffer from both ethical and hygienic drawbacks.[3-8] Therefore,VR simulation seems to be an appealing alternative as a training model.

Next to important logistic advantages, the virtual simulators offer automated scoring with numerous computer-based metrics. These could be used for assessment of procedural laparoscopic skills, replacing the laborious and subjective rating by experts during live or videotaped procedures.[9] However, to allow for valid assessment, these parameters must correctly reflect the actual operative skill of the trainee. This so-called concept of construct validity is present when the simulator is indeed able to discriminate between persons with known differences in performance level on the simulated skill, the laparoscopic cholecystectomy, in real life.[10,11]

The LAP Mentor (Simbionix, Cleveland) is a laparoscopic virtual simulator that, next to basic laparoscopic skills, suturing, and knot tying, provides a structured stepwise training program for the laparoscopic cholecystectomy. This training program consists of 4 cholecystectomy exercises, representing 4 distinct parts of the procedure to ensure a stepwise acquisition of the technique. They focus on the dissection of the Calot triangle, clipping and cutting of the cystic artery and duct, and the dissection of the gall bladder from the liver bed. Furthermore, 6 full cholecystectomy procedures, each with specific patient characteristics (i.e., short cystic duct and variations in cystic artery position), are provided. For every exercise and for the full procedures, several assessment parameters are measured simultaneously.[12,13] Construct validity of this procedural part of the LAP Mentor was addressed in only a single previous study.[13] Discriminative power between groups was found to be limited, and data clearly showed a lack of quality control. When aiming for the assessment of surgical skills, stronger validity evidence is needed. Therefore, the present study verifies the construct validity of the "cholecystectomy" module of the LAP Mentor virtual simulator.

MATERIAL AND METHODS

Subjects

For study purposes, 26 participants were recruited. Of these subjects, 5 were expert abdominal laparoscopists (>100 laparoscopic cholecystectomies performed), 11 were surgical residents in training (10-30 cholecystectomies performed), and 10 were nonsurgical residents in training (novices with minimal laparoscopic experience). All participating surgical residents were in their second or third year of training. All the novice participants had previously attended several human cholecystectomies. Therefore, adequate cognitive procedural input was guaranteed, and all had some experience with laparoscopic equipment. None of the subjects had previous experience with the LAP Mentor virtual simulator. The nature of the study was explained to all the subjects before enrollment, and informed consent was obtained from all the subjects.

LAP Mentor Virtual Simulator

The LAP Mentor is a computer-based VR simulator for learning laparoscopic skills, featuring 2 mock working instruments and a camera. Instrument and camera movements are translated into a virtual surgical environment, including haptic feedback, and displayed through a 17-in. flat liquid crystal display. The laparoscopic cholecystectomy is separated into 4 distinct parts of the cholecystectomy procedure (cholecystectomy exercises) to ensure a stepwise acquisition of the technique (Fig. 1). Furthermore, 6 full procedures are available of which only the first one was used for this study. Colored structures guide the trainee during the procedural exercise but are not provided during the full procedures. Diathermy, graspers, endoscissors, and a clip applier are available for use.[12,13]

Performance Evaluation

The quantitative parameters are measured in all exercises, as well as during the full procedures: total time needed to perform the exercise, the number of movements of the right and left hand, total path length of the right and left hand, and average speed of the right and left hand. For each of these motion parameters, a composite score for both hands (sum for "movements" and "path length"; average for "speed") was calculated so as to exclude the effect of dexterity (hand dominance) on performance scores.

The parameters assessing quality of performance are different for each exercise. All these parameters indicate better performance with a higher value. In the first exercise (Fig. 1A), the trainer provides the accuracy rate of clipping and cutting, that is, the percentage of clips and cutting maneuvers performed on the marked lines. In the second exercise (Fig. 1B), the trainer provides a safe clipping distance, that is, the distance between the proximal and the distal clip on the cystic artery and duct, and a safe cutting distance, that is, the distance between the division and the closest clip, either proximal or distal. A third distance that is measured by the trainer, between the distal clip and the infundibulum, seemed clinically irrelevant and was not used in this study. The safe clipping and cutting distance were summed for further calculations and called the safety parameter of clipping and cutting (measured in mm). In the third exercise (Fig. 1C), the trainer provides 4 independent quality parameters: the accuracy rate of dissection, that is, the percentage of time cautery is performed within the correct area (indicated with a blue color on the screen), which decreased when cautery is performed in the area of the common bile duct or hepatic artery; the efficiency rate of cautery, that is, the percentage of time cautery is applied in actual contact with the adhesions; the safety rate of cautery,

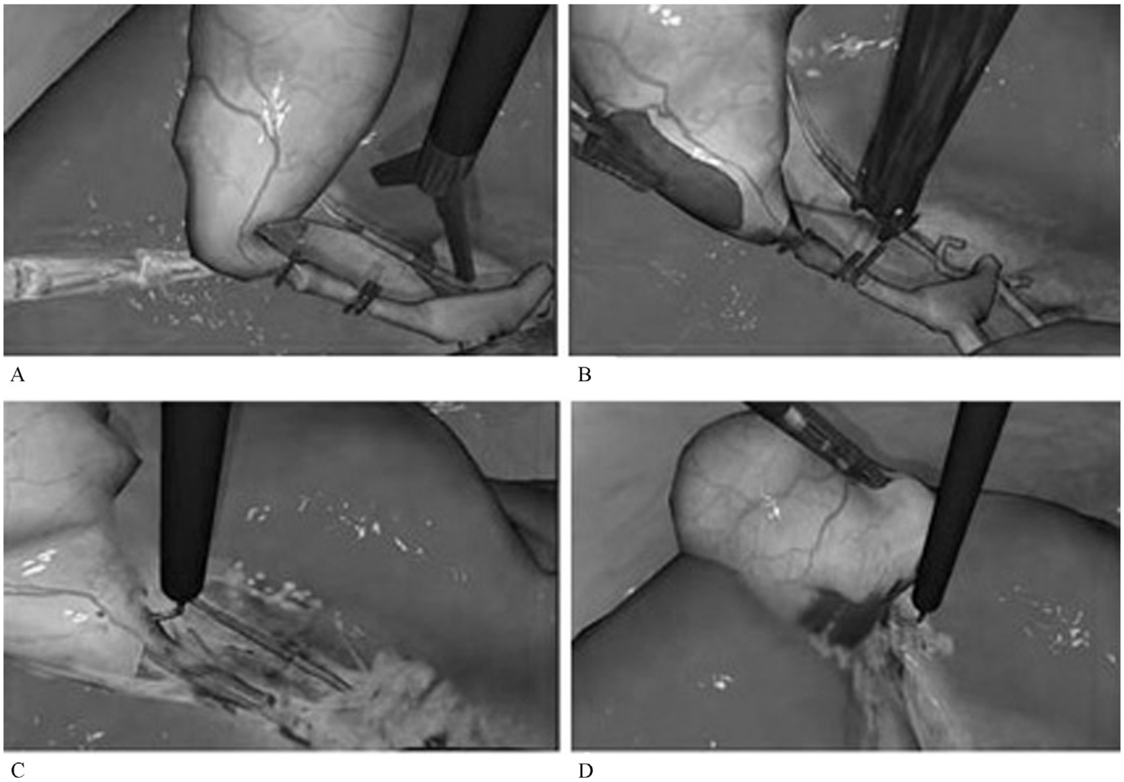

FIGURE 1. LAP Mentor cholecystectomy exercises. (A) *Exercise 1*: clipping and cutting on a retracted gall bladder. (B) *Exercise 2*: clipping and cutting while retracting the gall bladder. (C) *Exercise 3*: dissection of the Calot triangle. (D) *Exercise 4*: dissection of the gall bladder from the liver bed.

that is, the percentage of time cautery is applied at a sufficient distance (more than 5 mm) from the biliary system; and the completed dissection, that is, the percentage of the colored artery and duct that are exposed. In the fourth exercise (Fig. 1D), the trainer provides the accuracy rate of dissection, which is the percentage of the highlighted adhesions that are dissected, and the efficiency rate of cautery, which is the percentage of time cautery is applied in actual contact with the adhesions. Next to these quality parameters, the number of serious complications, as low as possible, is measured in each exercise (cautery or cutting the duct or artery when no clips are in position, cautery or cutting the common bile duct or hepatic artery, gall bladder perforations, and noncauterized bleeding). Because of the low incidence of these serious complications, these were summed over the 4 exercises. During the full procedures, the safe clipping and cutting distances are provided, which were again summed as described in exercise 2. Furthermore, safety and efficiency of cautery, as previously described, and the number of serious complications are measured. Accuracy rates or completed dissection is not provided here as there are no colored structures to guide the trainee.

Study Setup

All subjects attended a didactic hands-on session of all the 4 exercises, including a presentation of the parameters measured for assessment (familiarization run). Immediately afterwards, they performed each exercise twice and the first of 6 full procedures once. Exercises were excluded and redone if technical or software problems occurred. Performance parameters for the procedural exercises were averaged for the 2 attempts. The learning curve effect was minimized by the hands-on instruction session and by averaging the 2 trials for each exercise.

Statistical Analysis

Comparison of performance between the 3 groups can assess whether each parameter is construct valid. Performance was compared using the Kruskal-Wallis and Mann-Whitney *U* tests. Values were given as median (interquartile range), if not stated otherwise. This study was performed primarily as an exploratory study for the identification of valid outcome parameters. Therefore, no further corrections for multiple testing were made and $p < 0.05$ was considered statistically significant. Expert level was defined as the median score for each parameter measured over the 5 expert values.

RESULTS

Procedural Exercises

Results for quantitative parameters are shown in Table 1. The parameters "time" and "number of movements"

TABLE 1. Quantitative Parameters are Shown for All Exercises (Ex) and the Full Procedure (FP). Data are Shown as Median (Interquartile Range). p-Values of the Kruskal-Wallis Tests (kkw) and Mann-Whitney *U* Tests Between Experts (E), Surgical Residents (R), and Novices (N) are Shown. Statistically significant data ($p < 0.05$) are shown in bold

Parameter		Experts	Residents	Novices	kkw	E/R	E/N	R/N
Time (s)	Ex 1	68 (55-75)	73 (64-94)	120 (102-132)	**0.007**	0.51	**0.003**	**0.01**
	Ex 2	66 (60-79)	89 (73-107)	135 (112-159)	<**0.001**	0.09	**0.001**	<**0.001**
	Ex 3	165 (95-172)	335 (258-370)	445 (334-530)	**0.01**	**0.04**	**0.008**	0.07
	Ex 4	160 (136-168)	188 (171-305)	513 (463-613)	**0.001**	0.09	**0.001**	**0.001**
	FP	401 (346-421)	587 (502-673)	865 (766-991)	<**0.001**	**0.006**	**0.001**	<**0.001**
Movements (number)	Ex 1	39 (32-41)	36 (30-39)	70 (48-88)	**0.003**	0.83	**0.008**	**0.001**
	Ex 2	42 (38-42)	56 (39-58)	86 (69-96)	**0.001**	0.45	**0.005**	<**0.001**
	Ex 3	155 (92-191)	266 (191-345)	407 (322-482)	**0.008**	0.09	**0.008**	**0.02**
	Ex 4	171 (128-178)	250 (176-423)	594 (537-730)	**0.001**	0.05	**0.001**	**0.001**
	FP	356 (337-363)	560 (471-609)	955 (818-1055)	<**0.001**	**0.003**	**0.001**	<**0.001**
Path length (cm)	Ex 1	99 (95-116)	101 (94-135)	141 (108-166)	0.12	—	—	—
	Ex 2	107 (103-113)	123 (97-134)	150 (133-188)	**0.008**	0.44	**0.013**	**0.006**
	Ex 3	386 (191-403)	477 (388-731)	675 (525-992)	0.06	—	—	—
	Ex 4	317 (238-357)	560 (375-866)	1152 (986-1306)	**0.001**	**0.028**	**0.001**	**0.005**
	FP	796 (692-886)	1129 (1064-1388)	1905 (1709-2103)	<**0.001**	**0.01**	**0.001**	<**0.001**
Speed (m/s)	Ex 1	2.7 (2.7-2.7)	2.6 (2.4-3.2)	2.1 (2.0-2.5)	**0.04**	0.51	**0.02**	**0.04**
	Ex 2	3.2 (2.5-3.3)	2.6 (2.1-2.9)	2.2 (2.1-2.5)	0.22	—	—	—
	Ex 3	2.5 (1.9-2.5)	2.0 (1.9-2.3)	1.9 (1.7-2.0)	0.64	—	—	—
	Ex 4	1.9 (1.9-2.0)	2.3 (2.1-2.6)	2.1 (2.1-2.2)	0.10	—	—	—
	FP	2.5 (2.3-2.9)	2.7 (2.2-3.5)	2.2 (2.1-2.4)	0.31	—	—	—

showed significantly worse performance by the novices compared with the experts in every exercise. For the parameter "path length," the difference between groups was only significant in exercises 2 and 4. In none of the exercises, these parameters were able to discriminate between all the 3 groups. The interquartile range for these 3 parameters in all the 4 exercises decreased from the novices to the experts, indicating a more consistent performance among the experts. The parameter "speed" was only significant in exercise 1.

Quality of performance parameters are shown in Table 2. In exercise 3, the parameter "accuracy rate of dissection" showed significantly better performance by the experts vs the residents and novices, whereas for "completed dissection," residents significantly outperformed both the other groups (Fig. 2). None of the other quality parameters exhibited significant differences.

Full Procedure 1

The parameters "time," "number of movements," and "path length" showed significant differences between all 3 groups (Table 1, Fig. 3). Neither the parameter "speed" nor the qualitative parameters revealed any significant differences (Tables 1 and 2).

TABLE 2. Qualitative Parameters are Shown for All Exercises (Ex) and the Full Procedure (FP). Data are Shown as Median (Interquartile Range). p-Values of the Kruskal-Wallis tests (kkw) and Mann-Whitney *U* tests Between Experts (E), Surgical Residents (R), and Novices (N) are Shown. CD, completed dissection; SC, serious complications. Statistically significant data ($p < 0.05$) are shown in bold

Parameter		Experts	Residents	Novices	kkw	E/R	E/N	R/N
Accuracy (%)	Ex 1	84 (74-90)	75 (58-77)	80 (69-87)	0.34	—	—	—
Safety (mm)	Ex 2	7.5 (6.5-7.7)	7.7 (7.2-9.3)	8.6 (7.9-9.7)	0.39	—	—	—
Accuracy (%)	Ex 3	97 (95-99)	85 (73-88)	70 (67-77)	**0.009**	**0.02**	**0.003**	0.10
Efficiency (%)	Ex 3	72 (67-73)	58 (55-65)	67 (58-71)	0.05	—	—	—
Safety (%)	Ex 3	44 (39-51)	63 (52-70)	74 (45-80)	0.20	—	—	—
CD (%)	Ex 3	47 (45-70)	93 (71-98)	60 (50-66)	**0.01**	**0.03**	0.77	**0.004**
Accuracy (%)	Ex 4	45 (42-58)	44 (33-54)	59 (49-64)	0.07	—	—	—
Efficiency (%)	Ex 4	73 (65-75)	63 (54-75)	73 (61-75)	0.44	—	—	—
SC (number)	Ex	0.5 (0.5-1)	1.5 (0.3-2.8)	0.5 (0.0-1.6)	0.38	—	—	—
Safety clip/cut (mm)	FP	6.3 (5.6-6.9)	9.3 (5.4-10.8)	8.8 (6.9-12.4)	0.26	—	—	—
Efficiency cautery (%)	FP	61 (58-63)	66 (58-73)	68 (61-71)	0.46	—	—	—
Safety cautery (%)	FP	49 (43-52)	67 (49-77)	66 (63-70)	0.26	—	—	—
SC (number)	FP	1.0 (1.0-2.0)	2.0 (1.0-3.0)	4.5 (1.3-6.0)	0.18	—	—	—

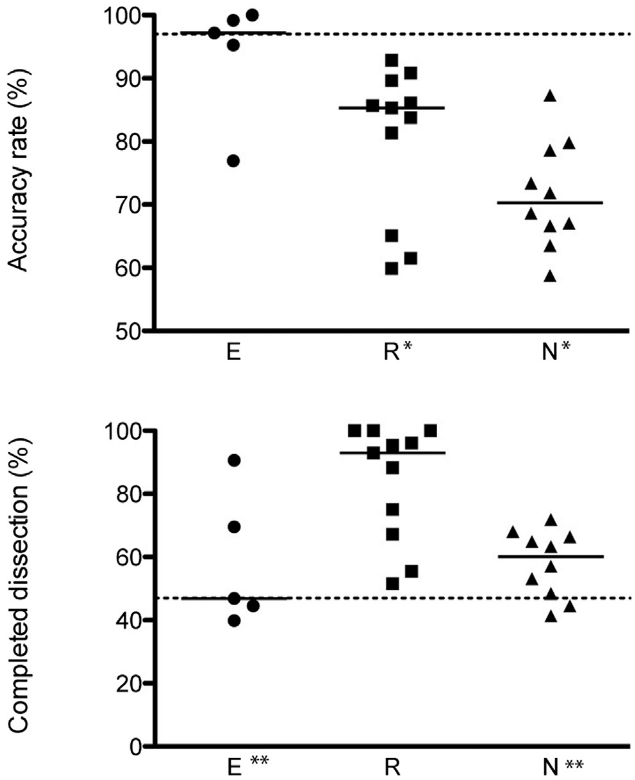

FIGURE 2. "Accuracy rate" and "completed dissection" for experts (E), surgical residents (R), and novices (N) in exercise 3. Data are shown as scatter dot plots with median indicated. * Significantly outperformed by experts and ** significantly outperformed by residents.

DISCUSSION

VR as a training model is an upcoming tool in surgical training programs. Initial research with virtual simulators has successfully focused on basic psychomotor and suturing skills programs,[14,15] but more robust and inexpensive video trainers are probably equally or even more effective and apparently more appealing and realistic to the trainees.[4-7] Therefore, the benefit of virtual simulation probably lies in the more advanced stages of training, where focus shifts toward procedural skills. Next to significant logistic advantages compared with live animal or cadaver organ training models, VR simulators provide automated scoring with numerous objective metrics that can be used for assessment. This would offer an appealing alternative for the laborious and subjective rating by the experts during live or videotaped procedures.[3,8,9] However, before integrating this tool into current practice, its ability to correctly discriminate between different levels of skill, the so-called construct validity, has to be proven.[10,11]

This study focused on the construct validation process of the LAP Mentor virtual simulator, which provides a stepwise training program for the laparoscopic cholecystectomy. Other virtual simulators similarly offer cholecystectomy training programs,[7,16] but the high degree of fidelity concerning surgical setting, mechanical interaction, and physiopathological behavior[7] and the unique combination of a stepwise cholecystectomy training program with several full procedures featuring specific anatomical variations and patient characteristics[16] seem to be important advantages of the LAP Mentor.

In this study, the quantitative parameters "time" and "number of movements" showed significant differences between experts and novices in all exercises. Similar results were noted for the parameter "path length" in exercises 2 and 4. Three groups of different performance levels were included to more accurately assess the discriminative power of the parameters, and it was only in the full procedure that the difference between all 3 groups was significant. Assessment of procedural skill could be used for high-stake

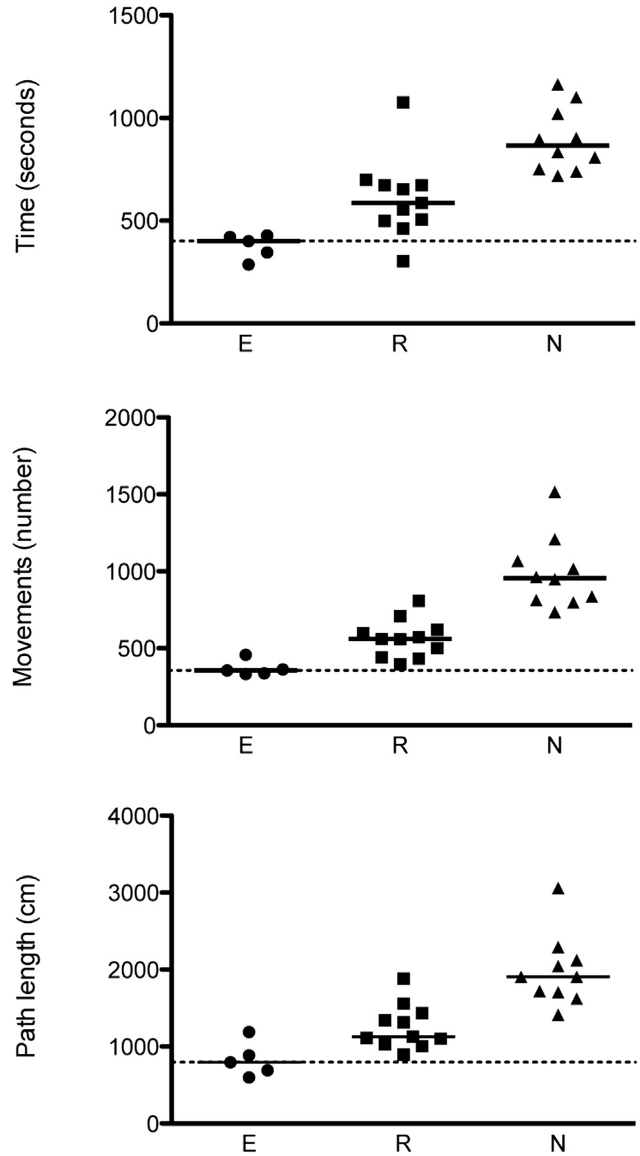

FIGURE 3. Significant quantitative parameters in the full procedure for (E) experts, (R) surgical residents, and (N) novices. Mann-Whitney *U* tests between all the 3 groups showed $p < 0.05$.

purposes, such as certification of surgeons or as a prerequisite for residents before operating on human beings. It could also have value to determine training outcome, for example, in the validation of virtual simulators as training tools, an application for which it is currently already in use.[17] For either purpose, a test is needed that is able to detect more subtle differences than those between the novices and the experts.[10] Therefore, we believe that only the full procedure, because of its higher discriminative power, has the potential to serve as an assessment tool. Results for the parameter speed were inconsistent and even more important; this parameter provides no specific suggestions toward better performance. When intentionally moving at higher speeds, more abrupt movements are likely to be performed, and when intentionally moving at lower speeds, the exercise is performed too slowly. We, therefore, consider this parameter unsuited for the assessment purposes.

Concerning the quality of performance, only accuracy rate in the dissection of the Calot triangle (exercise 3) showed significantly better performance by the experts. For the other parameters, such as completed dissection of the Calot triangle ($p < 0.05$) or all the safety scores ($p > 0.05$), the experts were even outperformed by less-experienced groups. None of the quality parameters in the full procedure proved to be construct valid. It is difficult to determine quality of performance as surgery is not a strictly mechanical process where every step is clearly defined. It is possible that experts scored lower on some items, because they are well aware just how far they can go, they walk the edges to optimize time vs quality. Conversely, even the parameter serious complications (i.e., perforation and bleeding) with obvious clinical relevance did not discriminate between groups either. Thus, at this moment, the virtual trainer lacks the realism that is necessary to mimic the possible pitfalls and complications during a laparoscopic cholecystectomy. In the evaluation of procedural skills, the quality of performance is of utmost importance because we need safe rather than fast surgeons. Furthermore, the pure technical, quantitative characteristics of trainee performance can easily be captured during simple psychomotor evaluation,[18] whereas it is the quality of tissue handling and dissection skills that constitute the surplus value of procedural skills assessment. We, therefore, conclude that the LAP Mentor procedural trainer is not (yet) suited for the assessment of procedural surgical skills.

Some limitations of the study include the small number of participants and thus power problem and the possibility of type II errors. Conversely, the discriminative ability of the variable is only meaningful when performance differences between these very distinct groups are large, diminishing the number of participants needed. It is noteworthy that because of the exploratory character of the study, no correction for multiple testing was made. Obviously, this prevents us from making strong claims about a specific variable. However, we do believe that the consistency of the results concerning the quantitative parameters "time," "movements," and "path length" indicates existing construct validity, whereas the complete lack of convincing results on the qualitative parameters makes further investigation indispensible. In addition, we did not assess other demographics of the 3 study groups, such as hand dominance or prior video gaming experience. The latter is known to influence simulator performance.[1] As this is expected to be higher in the novice group, this might have attributed to the fact that they outperformed the experts on some quality parameters. Finally, this study assessed all parameters that were provided by the manufacturer. As suggested by Korndorffer et al.[10] it is preferable to first identify which metrics are meaningful and relevant to surgical performance (content validation by surgical experts) before testing them for construct validity.

Previous research on the LAP Mentor virtual trainer mainly focused on its basic skills program,[17,19-22] whereas only a single study by Aggarwal et al.[13] specifically investigated the procedural tasks of the simulator. They identified the same weaknesses, namely a lack of difference between intermediate and expert subjects and accuracy rate of dissection in exercise 3, being the only qualitative parameter that proved construct validity. Other virtual trainer systems providing laparoscopic procedural exercises, for example, the LapSim cholecystectomy module[23,24] or the Mentice laparoscopic nephrectomy VR simulator,[25] show similar shortcomings, especially concerning quality control. However, although extensive validity evidence needs to be established for simulators to be used for assessment, the strict concept of validation does not apply to training.[10] What needs to be established is whether training on the simulator improves a subject's ability in the operating room through transferability studies. The parameters that are construct valid can be used to assess progress during training. For this purpose, the quantitative parameters seem useful. However, because of the current lack of construct validity for quality control, it is probably necessary to add external qualitative feedback by a surgical expert. Further study is needed to elucidate the role of the LAP Mentor virtual simulator in a training context.

CONCLUSION

Only the full procedure of the LAP Mentor procedural simulator has enough discriminative power to claim construct validity. However, the lack of quality control, which is indispensible in the evaluation of procedural skills, makes it currently unsuited for the assessment of procedural laparoscopic skills. The role of the simulator in a training context remains to be elucidated.

ACKNOWLEDGMENTS

We would like to thank the staff members of the Department of Abdominal Surgery of the University Hospital, Leuven and the residents in training for their participation in the study.

REFERENCES

1. Wanzel KR, Ward M, Reznick RK. Teaching the surgical craft: from selection to certification. *Curr Probl Surg*. 2002;39(6):573-659.
2. Korndorffer JR, Stefanidis D, Scott DJ. Laparoscopic skills laboratories: current assessment and a call for resident training standards. *Am J Surg*. 2006;191(1):17-22.
3. Berg DA, Milner RE, Fisher CA, Goldberg AF, Dempsey DT, Grewal H. A cost-effective approach to establishing a surgical skills laboratory. *Surgery*. 2007;142(5):712-721.
4. Hamilton EC, Scott DJ, Fleming JB, et al. Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. *Surg Endosc*. 2002;16(3):406-411.
5. Kanumuri P, Ganai S, Wohaibi EM, Bush RW, Grow DR, Seymour NE. Virtual reality and computer-enhanced training devices equally improve laparoscopic surgical skills in novices. *JSLS*. 2008;1(3):219-226.
6. Botden SM, Torab F, Buzink SN, Jakimowicz JJ. The importance of haptic feedback in laparoscopic suturing training and the additive value of virtual reality simulation. *Surg Endosc*. 2008;2(5):1214-1222.
7. Lamata P, Gómez EJ, Bello F, Kneebone RL, Aggarwal R, Lamata F. Conceptual framework for laparoscopic VR simulators. *IEEE Comput Graph Appl*. 2006;26(6):69-79.
8. Rusche B. The 3Rs and animal welfare—conflict or the way forward? *ALTEX*. 2003;20(suppl 1):63-76.
9. Vassiliou MC, Feldman LS, Fraser SA, et al. Evaluating intraoperative laparoscopic skill: direct observation versus blinded videotaped performances. *Surg Innov*. 2007;14(3):211-216.
10. Korndorffer JR Jr, Kasten SJ, Downing SM. A call for the utilization of consensus standards in the surgical education literature. *Am J Surg*. 2010;199(1):99-104.
11. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surg Endosc*. 2003;17(10):1525-1529.
12. Ayodeji ID, Schijven M, Jakimowicz J, Greve JW. Face validation of the Simbionix LAP Mentor virtual reality training module and its applicability in the surgical curriculum. *Surg Endosc*. 2007;21(9):1641-1649.
13. Aggarwal R, Crochet P, Dias A, Misra A, Ziprin P, Darzi A. Development of a virtual reality training curriculum for laparoscopic cholecystectomy. *Br J Surg*. 2009;96(9):1086-1093.
14. Carter FJ, Schijven MP, Aggarwal R, et al. Work Group for Evaluation and Implementation of Simulators and Skills Training Programmes. Consensus guidelines for validation of virtual reality surgical simulators. *Surg Endosc*. 2005;19(12):1523-1532.
15. Gurusamy K, Aggarwal R, Palanivelu L, Davidson BR. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. *Br J Surg*. 2008;95(9):1088-1097.
16. Schijven M, Jakimowicz JJ. Virtual reality surgical laparoscopic simulators. *Surg Endosc*. 2003;17(12):1943-1950.
17. Lucas S, Tuncel A, Bensalah K, et al. Virtual reality training improves simulated laparoscopic surgery performance in laparoscopy naïve medical students. *J Endourol*. 2008;22(5):1047-1051.
18. Korndorffer JR, Clayton JL, Tesfay ST, et al. Multicenter construct validity for southwestern laparoscopic videotrainer stations. *J Surg Res*. 2005;128(1):114-119.
19. McDougall EM, Corica FA, Boker JR, et al. Construct validity testing of a laparoscopic surgical simulator. *J Am Coll Surg*. 2006;202(5):779-787.
20. Zhang A, Hünerbein M, Dai Y, Schlag PM, Beller S. Construct validity testing of a laparoscopic surgery simulator (Lap Mentor): evaluation of surgical skills with a virtual laparoscopic simulator. *Surg Endosc*. 2008;22(6):1440-1444.
21. Andreatta PB, Woodrum DT, Birkmeyer JD, et al. Laparoscopic skills are improved with Lap Mentor training: results of a randomized, double blinded study. *Ann Surg*. 2006;243(6):854-863.

22. Lucas SM, Zeltser IS, Bensalah K, et al. Training on a virtual reality laparoscopic simulator improves performance of an unfamiliar live laparoscopic procedure. *J Urol*. 2008;180(6):2588-2591.

23. Aggarwal R, Grantcharov TP, Eriksen JR, et al. An evidence-based virtual reality training program for novice laparoscopic surgeons. *Ann Surg*. 2006;244(2):310-314.

24. Fairhurst K, Strickland A, Maddern G. The LapSim virtual reality simulator: promising but not yet proven. *Surg Endosc*. 2011;25(2):343-355.

25. Wijn RP, Persoon MC, Schout BM, Martens EJ, Scherpbier AJ, Hendrikx AJ. Virtual reality laparoscopic nephrectomy simulator is lacking in construct validity. *J Endourol*. 2010;24(1):117-122.